Projected Gradient Descent

CPSC 406 – Computational Optimization

Projected gradient descent

- projection onto a convex set
- gradient projection method

Orthogonal projection

For a set $C \subset \mathbb{R}^n$ closed convex, the **projection** of a point $x \in \mathbb{R}^n$ onto C is the point

$$\mathbf{proj}_C(x) = \underset{z \in C}{\operatorname{argmin}} \|x - z\|$$

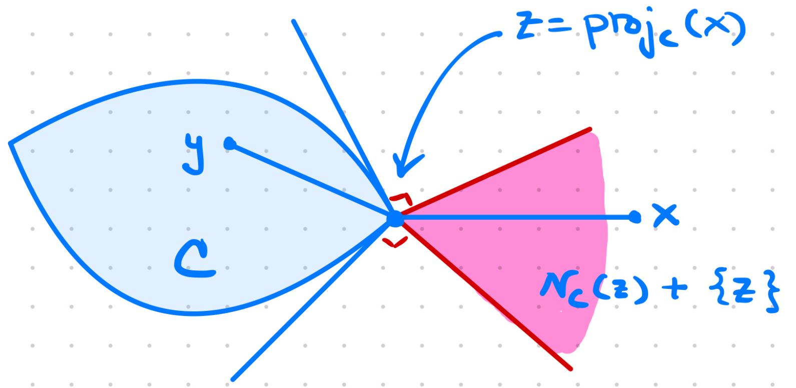
Properties:

1. if $x \in C$, then $\mathbf{proj}_C(x) = x$
2. $\mathbf{proj}_C(x)$ is unique (objective is strictly convex)
3. $z = \mathbf{proj}_C(x) \iff z \in C \quad \text{and} \quad (x-z)^T(y-z) \leq 0 \quad \forall y \in C$

proof of (3):

- Let $g(z) = \frac{1}{2}\|x - z\|^2$
- By optimality,

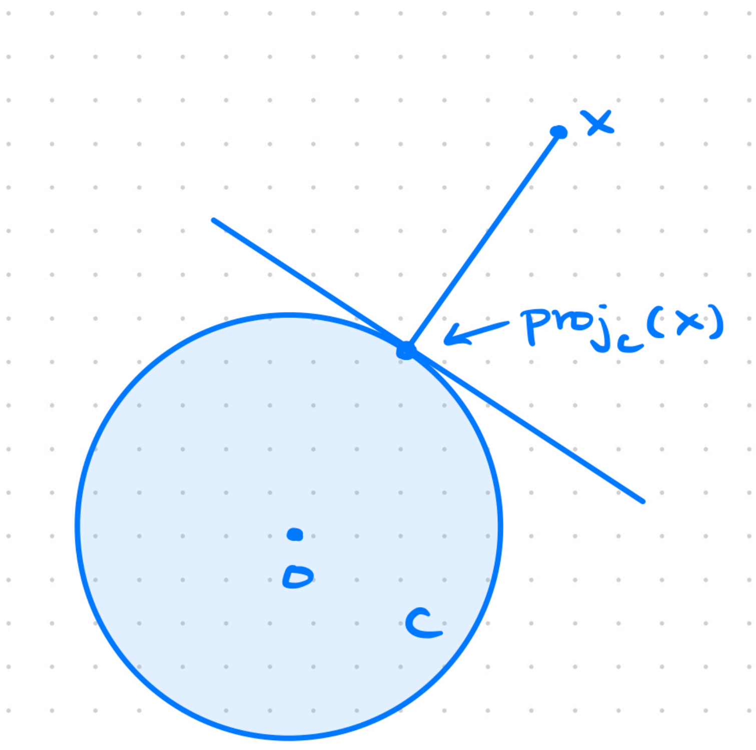
$$z = \mathbf{proj}_C(x) \iff -\nabla g(z) = x - z \in \mathcal{N}_C$$



Projection onto 2-norm ball

$$C = \{x \in \mathbb{R}^n \mid \|x\|_2 \le \alpha\} = \alpha\mathbb{B}_2 \quad (\alpha \ge 0)$$

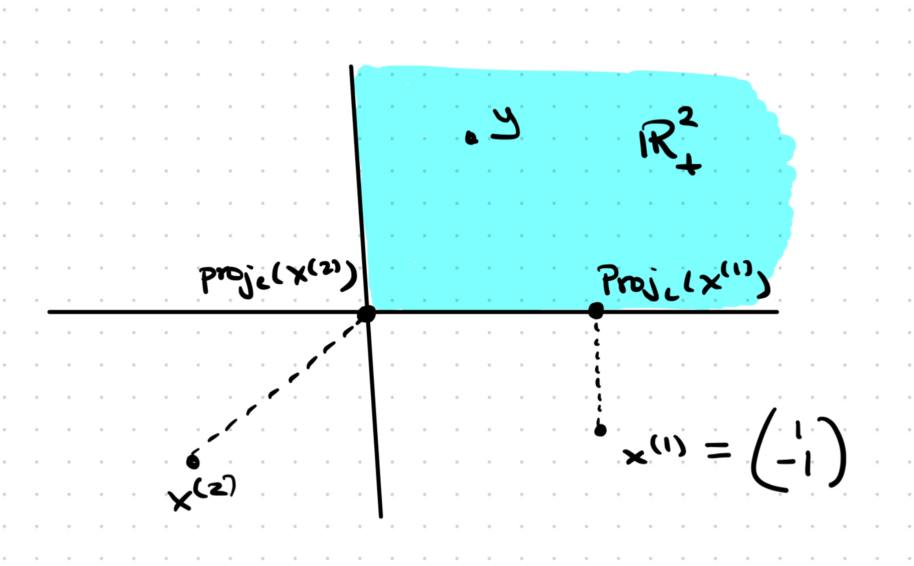
$$\mathbf{proj}_C(x) = \begin{cases} x & \text{if } \|x\|_2 \le \alpha \\ \alpha \frac{x}{\|x\|_2} & \text{if } \|x\|_2 > \alpha \end{cases}$$



Projection onto positive orthant

$$C = \mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\}$$

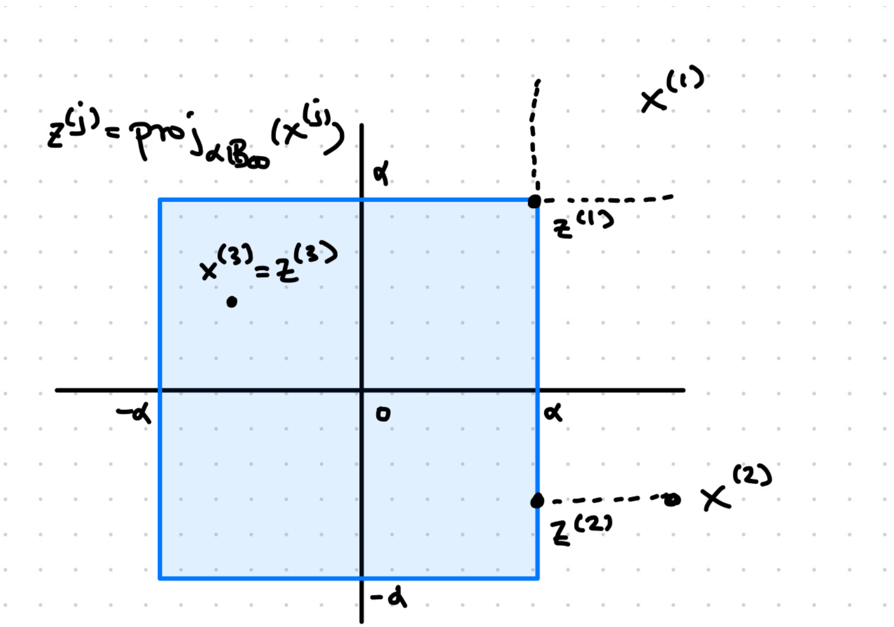
$$\mathbf{proj}_C(x) = \begin{bmatrix} \max\{0, x_1\} \\ \vdots \\ \max\{0, x_n\} \end{bmatrix}$$



Projection onto infinity-norm ball

$$C = \{x \in \mathbb{R}^n \mid \|x\|_\infty \le \alpha\} = \alpha\mathbb{B}_\infty \quad (\alpha \ge 0) \qquad \|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$$

$$\mathbf{proj}_C(x) = \begin{bmatrix} \mathbf{sign}(x_1) \cdot \min\{\alpha, |x_1|\} \\ \vdots \\ \mathbf{sign}(x_n) \cdot \min\{\alpha, |x_n|\} \end{bmatrix}$$



Question

Compute the projection of the vector $x = (1, 1, 1)$ onto the affine set

$$C = \{x \in \mathbb{R}^n \mid Ax = b\},$$

where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad b = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$$

a. $(-3.056, 0.11, 3.27)$
b. $(0.11, -3.056, 3.27)$
c. $(0.11, 3.27, -3.056)$
d. $(3.27, 0.11, -3.056)$

Example: de-biasing

$$C = \{x \in \mathbb{R}^n \mid e^T x = 0\}, \quad e = (1, 1, \ldots, 1), \quad e^T x = \sum_{i=1}^n x_i$$

By optimality conditions of the problem

$$\mathbf{proj}_C(x) = \text{argmin}\,\{\tfrac{1}{2}\|z - x\|^2 \mid e^T z = 0\}$$

we have

$$x - \mathbf{proj}_C(x) \in \mathbf{range}(e) = \alpha e \quad \text{for some } \alpha \in \mathbb{R}$$

To find α, premultiply by e^T:

$$e^T(x - \mathbf{proj}_C(x)) = \alpha e^T e = n \quad \Longrightarrow \quad \alpha = \frac{e^T x}{n} = \mathbf{avg}(x)$$

Thus,

$$\mathbf{proj}_C(x) = x - \frac{e^T x}{n} e$$

Projected gradient descent

$$\min_x \{f(x) \mid x \in C\}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ convex and smooth
- $C \subset \mathbb{R}^n$

Algorithm:

- Start from $x_0 \in C$
- For $k = 0, 1, 2, \ldots$
  - $g_k = \nabla f(x_k)$
  - linesearch on $\phi(\alpha) = f(\mathbf{proj}_C(x_k - \alpha g_k))$ (eg, backtracking, constant stepsize, etc.)
  - $x_{k+1} = \mathbf{proj}_C(x_k - \alpha_k g_k)$
  - stop if $\|x_{k+1} - x_k\|$ is small

Stationarity

$$x^* \in \operatorname*{argmin}_{x \in C} f(x) \quad \iff \quad x^* = \mathbf{proj}_C(x^* - \alpha \nabla f(x^*)) \quad \forall \alpha > 0$$

By projection theorem:

$$(x^* - \alpha \nabla f(x^*) - x^*)^T (z - x^*) \leq 0 \quad \forall z \in C$$

equivalently,

$$-\alpha \nabla f(x^*)^T (z - x^*) \leq 0 \quad \forall z \in C$$

Use definition of Normal cone to deduce equivalent condition:

$$-\nabla f(x^*) \in \mathcal{N}_C(x^*)$$